Synthetic Activators of Cell Migration Designed by Constructive Machine Learning

Bruns, Dominique ; Merk, Daniel ; Santhana Kumar, Karthiga ; Baumgartner, Martin ; Schneider, Gisbert

Abstract: Constructive machine learning aims to create examples from its learned domain which are likely to exhibit similar properties. Here, a recurrent neural network was trained with the chemical structures of known cell-migration modulators. This machine learning model was used to generate new molecules that mimic the training compounds. Two top-scoring designs were synthesized, and tested for functional activity in a phenotypic spheroid cell migration assay. These computationally generated small molecules significantly increased the migration of medulloblastoma cells. The results further corroborate the applicability of constructive machine learning to the de novo design of druglike molecules with desired properties.

DOI: https://doi.org/10.1002/open.201900222

Originally published at:
Bruns, Dominique; Merk, Daniel; Santhana Kumar, Karthiga; Baumgartner, Martin; Schneider, Gisbert (2019). Synthetic Activators of Cell Migration Designed by Constructive Machine Learning. ChemistryOpen, 8(10):1303-1308.
DOI: https://doi.org/10.1002/open.201900222

ChemPubSoc Europe

DOI: 10.1002/open.201900222

Open Access ChemistryOPEN Full Papers

Synthetic Activators of Cell Migration Designed by Constructive Machine Learning

Dominique Bruns,[a] Daniel Merk,[a] Karthiga Santhana Kumar,[b] Martin Baumgartner,*[b] and Gisbert Schneider*[a]

Constructive machine learning aims to create examples from its learned domain which are likely to exhibit similar properties. Here, a recurrent neural network was trained with the chemical structures of known cell-migration modulators. This machine learning model was used to generate new molecules that mimic the training compounds. Two top-scoring designs were synthesized, and tested for functional activity in a phenotypic spheroid cell migration assay. These computationally generated small molecules significantly increased the migration of medulloblastoma cells. The results further corroborate the applicability of constructive machine learning to the de novo design of druglike molecules with desired properties.

1. Introduction

The collection of methods enabling computational de novo design has recently been expanded by techniques utilizing machine learning, rather than explicit chemical transformations, for the construction of potentially novel molecules.[1,2] Certain classes of artificial neural networks, especially recurrent neural networks (RNN), are particularly suitable for this purpose.[3–6] These machine learning models are trained with collections of chemical structures, often represented as SMILES strings.[5] After successful training, an RNN can sample new SMILES strings from the learned training data distribution, such that the generated molecules possess features of the training compounds.

We previously applied an RNN with long short-term memory cells (LSTM[7]), which was trained with bioactive small molecules from ChEMBL,[8,9] and by fine-tuning with small sets of known bioactives (transfer learning), reoriented to generate target-specific compounds. Prospective applications demonstrated that nuclear receptor modulators and natural-product-inspired bioactive molecules which were synthetically accessible and confirmed as having in vitro activity against the desired targets, could be obtained using this approach.[10,11] Here, we apply this constructive machine learning approach to the computational design of cell migration modulators.

[a] Dr. D. Bruns, Dr. D. Merk, Prof. Dr. G. Schneider
ETH Zurich, Department of Chemistry and Applied Biosciences, Vladimir-Prelog-Weg 4, CH-8093 Zurich, Switzerland
E-mail: gisbert.schneider@pharma.ethz.ch
[b] Dr. K. Santhana Kumar, PD Dr. M. Baumgartner
Pediatric Neuro-Oncology Research Group, Department of Oncology, Children's Research Center, University Children's Hospital Zurich, Lengghalde 5, CH-8008 Zurich, Switzerland
E-mail: martin.baumgartner@kispi.uzh.ch

Supporting information for this article is available on the WWW under https://doi.org/10.1002/open.201900222

Small molecules affecting cellular migration and chemotaxis are of interest both as tool compounds for studying the metastatic behavior of malignant cells, and as potential drugs, *e.g.* for stem cell mobilization.[12] The often-complex biochemical networks controlling these processes are known to heavily involve the chemokine receptors.[13] These regulate a range of signaling pathways with G-protein or arrestin dependent branches, in a manner dependent on receptor polymerization state.[14,15] The G-protein coupled receptors (GPCRs) and their endogenous chemokine ligands are vital for immunoregulation and have also been linked to cancer pathology and prognosis.[16,17] The chemokine CXCL12, and its receptors CXCR4 and ACKR3, plays a central role in the metastasis of malignant cells.[18,19] CXCR4 antagonists have been developed as cell migration modulators, including the approved drug Plerixafor (AMD3100) for stem cell mobilization.[12]

Here, we use a constructive RNN model[5] to generate small molecule cell migration inducers. A pre-trained RNN was fine-tuned with a set of known chemokine receptor ligands. The candidate molecules suggested by the constructive machine learning approach were prioritized based on their pharmacophore similarity to the fine-tuning set, and two of the de novo designs were synthesized. These compounds induced pronounced medulloblastoma cell migration in a phenotypic spheroid invasion assay (SIA).[20] RNAi mediated knockdown of CXCR4 abrogated the migration-inducing effects, indicating the relevance of this chemokine receptor for the observed phenotypic activity. These results corroborate the applicability of constructive machine intelligence to the task of generating new molecules with desired biological activity.

2. Results and Discussion

For computational de novo design of cell migration inducers we employed a pre-trained RNN model that had successfully been used for generating bioactive new chemical entities in previous studies.[10,11] This model was fine-tuned by transfer-learning using a collection of compounds exhibiting CXCR4

antagonistic activity in a variety of in vitro test systems with a potency threshold of <1 μM ("template" collection). As the RNN was pre-trained on SMILES strings with a maximum length of 74 characters, this cutoff was also applied to the template collection (Figure S1). To promote structural diversity in the computer-generated molecular structures, the templates were selected to avoid bias towards certain scaffolds. For the most frequently occurring atom scaffolds (Murcko scaffolds)[21] (where the scaffolds were contained in at least eight of the known CXCR4 antagonists), the three most active molecules were selected for RNN fine-tuning, and from the scaffolds with two to eight representatives, the two most active compounds were included. The fine-tuning set contained 25 structurally diverse molecules with annotated CXCR4 antagonistic bioactivity[22–30] (Table S1). Of these 25 molecules, two had previously been assessed for CXCL12-induced cell migration inhibition, one of which had an activity (IC_{50}) of 58 nM.[22]

After transfer learning, a total of 1000 SMILES strings were sampled from the fine-tuned RNN model, of which 911 were chemically valid and 792 were unique. To prioritize compounds for synthesis and in vitro pharmacological evaluation, we analyzed the pharmacophore similarity of the designs and the templates using the Chemically Advanced Template Search (CATS) descriptors.[31] CATS descriptors represent chemical structures by means of topologically cross-correlated pharmacophore features.[31] The de novo molecular designs were ranked according to their decreasing average distance to the 25 template molecules used for model fine-tuning. Compounds **1** and **2** were selected for synthesis based on their ranking position and building block availability.

Compound **1** was prepared from the commercially available building block **3** in a linear three-step strategy with 68% overall yield (Figure 1). A two-step Staudinger reaction involving the formation of azide **4** from alcohol **3** followed by reduction with triphenylphosphine/water afforded intermediate tetrahydroisoquinoline-8-amine (**5**). Subsequent reaction of **3** with 2-phenylacetyl chloride (**6**) yielded compound **1**. Since the generative model was not trained to capture stereochemical information, this aspect was also not considered in the synthesis of **1**. Compound **2** was synthesized from isocyanate **7** and amine **8** under microwave irradiation in a single step (Figure 1).

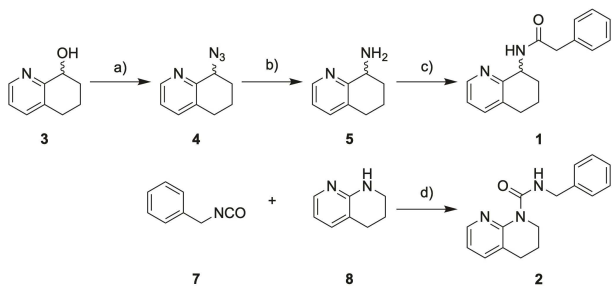

Figure 1. Synthesis of computer-generated designs **1** and **2**. *Reagents & conditions*: a) MsCl, NaN_3, 4-DMAP, CH_2Cl_2, DMF, $-4\,^{\circ}C\rightarrow$rt, 4 h, 82%; b) PPh_3, H_2O, THF, rt, 16 h, 90%; c) 2-phenylacetyl chloride (**6**), 4-DMAP, THF, $0\,^{\circ}C\rightarrow$rt, 60 min., 92%. d) neat, μw 80 °C, 1 h, 75%.

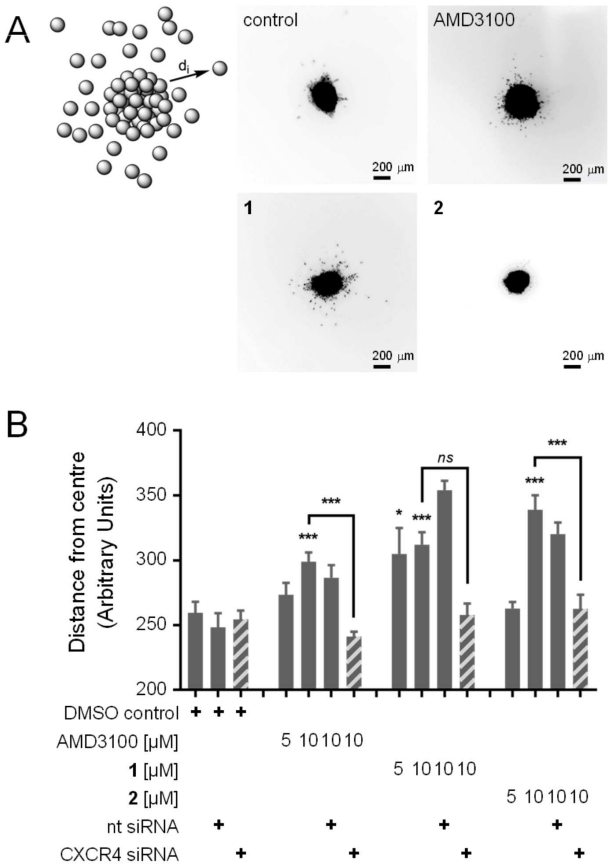

Figure 2. Phenotypic spheroid invasion assay (SIA). (a) Examples of microscopy images showing the effects of compounds **1** and **2**, the untreated control, and AMD3100. (b) Activity of computational designs **1** and **2** in the SIA. CXCR4 antagonist AMD3100 for comparison. Cell migration was induced by compounds **1**, **2**, and the known CXCR4 antagonist AMD3100. This effect was abrogated by siRNA mediated CXCR4 silencing. Non-targeting (nt) siRNA had no effect. Results are given as *mean* ± S.E.M.; $N=3$ independent repetitions with $n=3$ technical replicates each; $^{*}p<0.05$, $^{**}p<0.01$, $^{***}p<0.001$ (one-way ANOVA with Kruskal-Wallis test).

To study the effects of compounds **1** and **2** on cell migration, we employed a phenotypic spheroid invasion assay[20] (Figure 2) with a medulloblastoma cell line (DAOY) expressing CXCR4 (Figure S2). In this system, the CXCR4 antagonist AMD3100 induced significant cell migration which was abrogated by siRNA-mediated silencing of the GPCR. Non-targeting siRNA had no effect. These results suggest a CXCR4-dependent pro-migratory function of AMD3100. Apparently, this effect of AMD3100 is phenocopied when compounds **1** or **2** are applied. Both compounds dose-dependently induced cell migration with greater efficacy than AMD3100 (Figure 2b).

Silencing of CXCR4 by RNAi abolished the activity of AMD3100 as well as designs **1** and **2** (Figure S3), thereby confirming the involvement of this chemokine receptor in the observed effects. Closer inspection of cell migration activators **1** and **2** in a functional assay for CXCR4-mediated effects on cAMP levels revealed no activity at 50 μM, thus not showing AMD3100's inhibitory effect on cAMP production. Of note, the RNN fine-tuning set contained compounds that induce calcium mobilization through CXCR4, and exert anti-HIV activity. Our

in vitro tests, in contrast, were focused on compound effects on intracellular cAMP levels, and on cell mobility. The discrepancy between the observed readouts suggests that, at least partially, other, CXCR4-dependent pathways are responsible for the observed effects of cell migration activators **1** and **2**.

In an attempt to assess the potential co-involvement of other target proteins in the observed effects of **1** and **2** on cell migration, we predicted off-targets for both compounds with the SPiDER[32] and TIGER[33] software programs. These computational tools employ pre-trained self-organizing maps to identify pharmacophore similarities in query compounds with known reference ligands of pharmacological targets.[33] The predicted targets were assessed for their potential involvement in cell migration or chemotaxis according to the respective Gene Ontology[34,35] definitions. Positively predicted targets with known relevance in cell migration, and which are expressed in the DAOY cells used in the phenotypic assay, were selected for the further in vitro characterization of compounds **1** and **2** (Figure 3, Table S2). While no relevant activities ($\geq$25% effect at 50 μM) were observed for compound **1**, compound **2** slightly modulated the GPCRs somatostatin sst1, dopamine receptor D2S and chemokine receptor CCR10. The most prominent activity of **2** (partial agonism) was observed for the D2S receptor. This preliminary observation is in agreement with the known importance of dopamine D2 receptors in cancer cell migration and invasion[36]. This working hypothesis is further supported by target predictions with TIGER software for the RNN fine-tuning set of compounds. Three (12%) of these CXCR4 modulators were also predicted to interact with dopamine receptors, specifically with D2 and D3 subtypes.

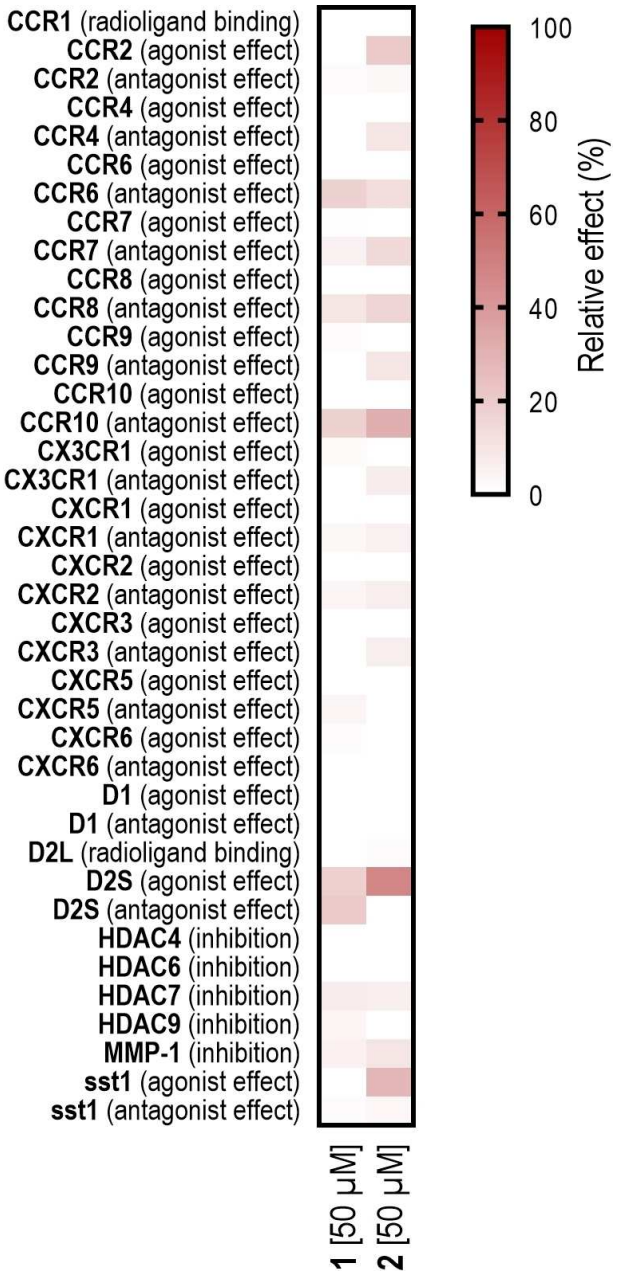

Figure 3. Functional profiling of compounds **1** and **2** on proteins involved in the regulation of cell migration. Data are presented as the mean of two technical replicates. Activities exceeding 25% were considered relevant. Compound **2** activated dopamine D2S receptor (47%), antagonized CCR10 (31%) and activated sst1 receptor (28%). No relevant activities were observed for compound **1**. Individual activity values are presented in Table S2.

3. Conclusions

The neural network employed for compound design was able to capture activity-relevant features from a collection of template compounds and autonomously construct new molecules with the desired bioactivity. The computer-generated compounds **1** and **2** were synthetically accessible with standard chemistry in rapid procedures and with high yields. Both tested compounds achieved the desired phenotypic effect and provide starting points for the systematic development of synthetic cell-migration modulators. Thus, in addition to generating molecules with isofunctional activity on a well-defined protein target, as shown previously, the machine learning approach was capable of generating de novo designs with isofunctional phenotypic effects. This result further corroborates the potential of constructive machine learning to provide innovative chemical hypotheses for drug discovery.

Experimental Section

Technical implementation. The RNN model was implemented in Python (3.6.1), using Tensorflow (1.4.1) and Keras (2.0.8) libraries. The generative runs were executed on a medium-grade Linux Workstation (Ubuntu 16.04) equipped with Intel(R) Core(TM) i7-6850 K CPU (3.60GHz, 32Gb DDR4 RAM), and a NVIDIA 1080 Ti 12GB (ASUS ROG) consumer-grade GPU. All further calculations were performed on Macintosh Workstation (OS X Yosemite, 10.10.5, 2× 2.26 GHz Quad-Core Intel Xeon with 48GB memory). MOE 2016.08 (The Chemical Computing Group, Montreal, Canada) was used for the standardization of molecules. Murcko scaffolds[21] were described using Python (3.5.3) with RDKit (2016.03.4), seaborn (0.7.1), matplotlib (2.0.0) and pandas (0.19.3) libraries.

Recurrent neural network. The general model structure and model training was performed as described in our earlier work.[5] Fine-tuning of the model was performed for five epochs using the

template collection of CXCR4 ligands. Stereochemical information was not taken into account during RNN training and fine-tuning.

Training data. The ChEMBL dataset used for the RNN training is described in ref. [5]. For model fine-tuning, CXCR4 ligands were selected from several sources.[8,9,22–30,37–41] Compounds were considered if they were reported as active in a functional assay with a potency <1 μM. In case several activities for one compound were reported, the strongest activity was kept. The molecular structures were canonicalized using MOE2016.

CATS descriptor. Chemically advanced template search (CATS) is a 2D descriptor encoding pharmacophoric features dependent on topological distances.[31] All computationally generated molecules and scaffolds were standardized to be neutral. CATS descriptors[31] were calculated using in-house software (SpeedCATS, implemented as Knime (2.12.02) node) with default parameterization (distance = 10, scaling = 'types').

Scaffold analysis. Molecules were neutralized using MOE 2016.08. Murcko decomposition was performed, keeping the atomic information but generalizing the compound to a framework[21]. The RDKit library function MurckoScaffold.GetScaffoldForMol() in python3 was used.

SPiDER target prediction. The molecules were standardized to nominal pH 7 (MOE 2016.08). Molecules supplied in SMILES format were converted to molecules using the 'Mol From Smiles' node, if molecules were stored in SDF format, they were loaded directly. SpeedCats software was used for the CATS descriptor[32] calculation and joined in a vector. SpeedCats settings were applied as for the CATS calculations. The respective MOE node was used to calculate the structural descriptors. The MOE-descriptors were stored in a vector as well and both descriptor vectors selected in the SPIDER node to calculate the predictions. Predictions with a p-value <0.05 were considered as positive.

TIGER target prediction. Target predictions were performed as described previously.[42] Compound structures were neutralized for computational analysis. Targets predicted with TIGER scores >1 were considered meaningful and kept for further analysis.

Chemical synthesis and analytics. All chemicals and solvents were reagent grade and used without further purification, unless specified otherwise. All reactions were conducted in oven-dried glassware under argon-atmosphere and in absolute solvents. NMR spectra were recorded on a Bruker AV 400 spectrometer (Bruker Corporation, Billerica, MA, USA). Chemical shifts (δ) are reported in ppm relative to TMS as reference; approximate coupling constants (J) are given in Hertz (Hz). Mass spectra were obtained on an Advion expression CMS (Advion, Ithaka, NY, USA) equipped with an Advion plate express TLC extractor (Advion) using electrospray ionization (ESI). High-resolution mass spectra were recorded on a Bruker maXis ESI–Qq-TOF-MS instrument (Bruker). Compound purity was analyzed by HPLC on a VWR LaChrom ULTRA HPLC (VWR, Radnor, PA, USA) equipped with a MN EC150/3 NUCLEODUR C18 HTec 5 μ column (Machery-Nagel, Düren, Germany) using a gradient (gradient 1: H_2O/MeCN 95:5 + 0.1 % formic acid isocratic for 5 min to H_2O/MeCN 5:95 + 0.1 % formic acid after additional 25 min and H_2O/MeCN 5:95 + 0.1 % formic acid isocratic for additional 5 min; gradient 2: H_2O/MeCN 70:30 + 0.1 % formic acid isocratic for 3 min to H_2O/MeCN 5:95 + 0.1 % formic acid after additional 9 min H_2O/MeCN 5:95 + 0.1 % formic acid isocratic for additional 2 min) at a flow rate of 0.5 ml min^{-1}, temperature of 27 °C and UV-detection at 245 nm and 280 nm. All final compounds for biological evaluation had a purity >95 % (area-under-the-curve for UV_{245} and UV_{280} peaks).

8-Azido-5,6,7,8-tetrahydroisoquinoline (4)

5,6,7,8-Tetrahydroquinoline-8-ol (**3**, 746 mg, 5.00 mmol, 1.00 eq) was dissolved in a mixture of methylene chloride (abs., 45 ml) and DMF (abs., 5 ml), and 4-DMAP (1.22 g, 10.0 mmol, 2.00 eq.) and sodium azide (975 mg, 15.0 mmol, 3.00 eq) were added. The mixture was cooled to −4 °C and methanesulfonyl chloride (780 μl, 1.15 g, 10.0 mmol, 2.00 eq) was added slowly. The mixture was allowed to warm to room temperature and stirred for 4 h. Sodium hydroxide solution (1 M, 100 ml) and hexane (100 ml) were then added, phases were separated and the aqueous layer was extracted with hexane (2 × 100 ml). The combined organic layers were washed with brine (200 ml), dried over magnesium sulfate, and the solvents were evaporated in vacuum. The crude product was purified by column chromatography using methylene chloride/methanol (97:3) as mobile phase to obtain the title compound as yellow oil (715 mg, 82 %). ^{1}H NMR (400 MHz, Chloroform-*d*) δ = 1.69–2.06 (m, 2H), 2.07–2.23 (m, 2H), 2.27–2.38 (m, 1H), 2.66–2.87 (m, 2H), 7.08 (dd, J = 7.8, 4.7 Hz, 1H), 7.33–7.40 (m, 1H), 8.42 (ddd, J = 4.7, 1.8, 0.9 Hz, 1H) ppm. ^{13}C NMR (101 MHz, Chloroform-*d*) δ = 17.45, 28.02, 32.47, 58.98, 123.21, 132.15, 137.55, 147.85, 154.52 ppm. MS (ESI+): *m/z* 175.4 ([M + H]$^+$).

5,6,7,8-Tetrahydroquinoline-8-amine (5)

8-Azido-5,6,7,8-tetrahydroisoquinoline (**4**, 696 mg, 4.00 mmol, 1.00 eq) and triphenylphosphine (315 mg, 5.00 mmol, 1.20 eq) were dissolved in THF (40 ml) and water (4 ml) was added. The mixture was stirred at room temperature for 16 h. The solvents were then evaporated in vacuum and the crude product was purified by column chromatography using methylene chloride/methanol (95:5) as mobile phase to obtain the title compound as yellow oil (533 mg, 90 %). ^{1}H NMR (400 MHz, Chloroform-*d*) δ = 1.59–1.78 (m, 2H), 1.85–1.94 (m, 1H), 2.09–2.20 (m, 1H), 2.63–2.81 (m, 2H), 3.95 (dd, J = 7.7, 5.4, 1H), 7.00 (ddd, J = 7.7, 4.8, 0.7, 1H), 7.30 (ddt, J = 7.8, 1.9, 1.0, 1H), 8.34 (ddt, J = 4.7, 1.7, 0.8, 1H) ppm. ^{13}C NMR (101 MHz, Chloroform-*d*) δ = 29.04, 31.96, 39.07, 51.44, 121.76, 129.17, 131.64, 136.80, 149.22 ppm. MS (ESI+): *m/z* 149.1 ([M + H]$^+$).

2-Phenyl-*N*-(5,6,7,8-tetrahydroquinoline-8-yl)acetamide (1)

5,6,7,8-Tetrahydroquinoline-8-amine (**4**, 38 mg, 0.25 mmol, 1.00 eq) was dissolved in THF (abs., 5 ml), 4-DMAP (61 mg, 0.50 mmol, 2.00 eq) was added, the mixture was cooled to 0 °C and 2-phenylacetylchloride (**5**, 43 μl, 50 mg, 0.33 mmol, 1.30 eq.) was added dropwise under stirring. The mixture was allowed to warm and stirred for 60 min at room temperature. Sodium hydroxide solution (1 M, 5 ml) and ethyl acetate (5 ml) were added, phases were separated and the aqueous layer was washed with brine (10 ml). The organic layer was dried over magnesium sulfate and the solvents were evaporated in vacuum. The crude product was purified by column chromatography using methylene chloride/methanol (97:3) as mobile phase and subsequently washed with boiling hexane to obtain the title compound as colorless solid (61 mg, 92 %). ^{1}H NMR (400 MHz, DMSO-d_6) δ = 1.69–1.99 (m, 4H), 2.67–2.85 (m, 2H), 3.43 (d, J = 14.0 Hz, 1H), 3.47 (d, J = 14.0 Hz, 1H), 4.90 (q, J = 6.0 Hz, 1H), 7.18–7.26 (m, 2H), 7.27–7.33 (m, 4H), 7.54 (dd, J = 7.7, 1.7 Hz, 1H), 8.36–8.46 (m, 2H) ppm. ^{13}C NMR (101 MHz, DMSO-d_6) δ = 19.31, 28.43, 30.14, 42.79, 49.38, 122.77, 126.65, 128.58, 129.44, 133.26, 137.19, 137.30, 147.50, 155.99, 169.61 ppm. MS(ESI+) *m/z* 267.0 ([M + H]$^+$). HRMS(ESI+) *m/z* calculated 267.1492 for $C_{17}H_{19}N_2O$ found 267.1486 ([M + H]$^+$). HPLC (gradient 1): RT 12.79 min.

N-Benzyl-3,4-dihydro-1,8-naphthyridine-1(2H)-carboxamide (2)

1,2,3,4-Tetrahydro-1,8-naphthyridine (**8**, 50 mg, 0.37 mmol; 1.0 eq.) and (isocyanatomethyl)benzene (**7**, 0.5 ml, 463 mg, 3.5 mmol; 10 eq.) were mixed in a 5 ml microwave vial. The mixture was irradiated for 5 minutes at 50 °C, followed by 1 hour at 80 °C. 1 M sodium hydroxide solution and ethyl acetate were then added, phases were separated, and the aqueous layer was extracted twice with ethyl acetate. The combined organic layers were dried with $MgSO_4$ and concentrated in vacuum. The crude product was purified by column chromatography using hexane/ethyl acetate (6:4) as mobile phase to obtain the title compound as colorless oil (74 mg, 75%). ^{1}H NMR (400 MHz, DMSO-d_6) δ = 1.87–1.78 (m, 2H), 2.79 (t, J = 6.3 Hz, 2H), 3.92–3.83 (m, 2H), 4.46 (d, J = 5.8 Hz, 2H), 6.96 (dd, J = 7.4, 4.9 Hz, 1H), 7.28–7.18 (m, 1H), 7.37–7.28 (m, 4H), 7.60–7.53 (m, 1H), 8.13–8.05 (m, 1H), 10.90 (t, J = 5.9 Hz, 1H) ppm.^{13}C NMR (101 MHz, DMSO-d_6) δ = 20.94, 27.51, 43.26, 43.45, 116.94, 121.89, 126.65, 126.98, 128.33, 138.32, 140.10, 143.41, 152.17, 155.77 ppm. HRMS(ESI+) *m/z* calculated 268.1444 for $C_{17}H_{19}N_2O$ found 268.1441 ($[M+H]^+$). HPLC (gradient 2): RT 11.55 min.

Spheroid invasion assay. The spheroid invasion assay (SIA) using the medulloblastoma tumour cell line DAOY was performed as described previously[20] to observe effects of **1** and **2** on cell migration. DAOY cells stably expressing lifeact (LA) Enhanced green fluorescent protein (EGFP) produced by lentiviral transduction with pLenti-LA-EGFP were used for SIA. In brief, 1000 DAOY LA-EGPF cells per 100 µl per well were seeded in a 96 well cell-repellent 96 well microplate (650790, Greiner Bio-one). The cells were incubated overnight at 37 °C to form spheroids. 70 µl of the medium was removed from each well, and remaining medium with spheroid was overlaid with 2.5% (final concentration) of ice cold bovine collagen 1 (5005-B, Advanced BioMatrix, San Diego, CA, USA). The collagen is allowed to polymerize for one hour at 37 °C. Following the polymerization of collagen, fresh serum free medium was added to the cells and then treated with 10 µM (final concentration) of the compounds. The embedded cells were allowed to invade the 3D collagen matrix for 24 hours, after which they were fixed with 4% PFA and stained with Hoechst. Images were acquired on an Axio Observer 2 mot plus fluorescence microscope (Zeiss, Munich, Germany) using a 5x objective. The extent of cell invasion was determined as the average of the distance invaded by the cells from the center of the spheroid as using automated cell dissemination counter (aCDc). D'Agostino and Pearson normality test followed by unpaired t-test were performed using GraphPad Prism version 7.00 for Apple (GraphPad Software, La Jolla California USA).

Functional cAMP assay. Compounds were tested in a single-concentration assay at 50 µM. CXCR4 (86-0007P), were assayed in functional assays testing their agonistic and antagonistic activity respectively measuring cAMP levels. Assays were provided at a fee-for-service basis by DiscoverX (Fremont, CA, USA).

Target panel screening. Compounds were tested in a single-concentration assay at 50 µM. CCR2 (2497, 2501), CCR4 (4229, 4230), CCR6 (4231, 4232), CCR7 (4233, 4234), CCR8 (4235, 4236), CCR9 (4237, 4238), CCR10 (4227, 4228), CX3CR1 (4239, 4240), CXCR1 (4241, 4242), CXCR2 (4243, 4244), CXCR3 (4245, 4246), CXCR5 (4247, 4248), CXCR6 (4249, 4250), D1 (1685, 1686), D2 S (2566, 2569), sst1 (2253, 2254) were assayed in functional assays testing their agonistic and antagonistic activity respectively. CCR1 (361), and D2 L (1405) were assayed in binding assays using scintillation counting. MMP-1 (510), HDAC4 (2493), HDAC6 (2495), HDAC7 (2610), and HDAC9 (2611) were assayed in enzyme-based assays using fluorimetry. Assays were provided at a fee-for-service basis by Cerep (Celle l'Evescault, France).

Acknowledgements

The authors thank Lukas Friedrich, Berend Huisman, Claudia Neuhaus and Sarah Haller for technical support. The authors thank Ryan Byrne for proof reading of the manuscript. This research was financially supported by the Swiss National Science Foundation (grant no. CR3212_159737 to G.S.), the Novartis Forschungsstiftung (FreeNovation: AI in Drug Discovery), and the ETH RETHINK initiative. inSili.com GmbH kindly provided access to the TIGER software.

Conflict of Interest

The authors declare the following competing interests: G.S. declares a potential financial conflict of interest in his role as life science industry consultants and cofounder of inSili.com GmbH, Zurich. D.M. was financially supported by an ETH Zurich Postdoctoral Fellowship (grant no. 16–2 FEL-07).

Keywords: chemoinformatics · chemotaxis · drug discovery · neural networks · phenotypic screening

[1] G. Schneider, *Nature Mach. Intell.* **2019**, *1*, 128–130.
[2] G. Hessler, K. H. Baringhaus, *Molecules* **2018**, *23*, 2520.
[3] a) G. Schneider, D. Clark, *Angew. Chem.* **2019**, *131*, 10906–10917; b) G. Schneider, D. Clark, *Angew. Chem. Int. Ed.* **2019**, *58*, 10792–10803.
[4] M. Olivecrona, T. Blaschke, O. Engkvist, H. Chen, *J. Cheminform.* **2017**, *9*, 48.
[5] A. Gupta, A. T. Mueller, B. J. H. Huisman, J. A. Fuchs, P. Schneider, G. Schneider, *Mol. Inf.* **2018**, *37*, 1700111.
[6] M. H. S. Segler, T. Kogej, C. Tyrchan, M. Waller, *ACS Cent. Sci.* **2018**, *4*, 120–131.
[7] S. Hochreiter, J. Schmidhuber, *Neural Comput.* **1997**, *9*, 1735–1780.
[8] A. Gaulton, L. J. Bellis, A. P. Bento, J. Chambers, M. Davies, A. Hersey, Y. Light, S. McGlinchey, D. Michalovich, B. Al-Lazikani, J. P. Overington, *Nucleic Acids Res.* **2012**, *40*, D1100–D1107.
[9] A. P. Bento, A. Gaulton, A. Hersey, L. J. Bellis, J. Chambers, M. Davies, F. A. Krüger, Y. Light, L. Mak, S. McGlinchey, M. Nowotka, G. Papadatos, R. Santos, J. P. Overington, *Nucleic Acids Res.* **2014**, *42*, D1083–D1090.
[10] D. Merk, F. Grisoni, L. Friedrich, G. Schneider, *Commun. Chem.* **2018**, *1*, 68.
[11] D. Merk, L. Friedrich, F. Grisoni, G. Schneider, *Mol. Inf.* **2018**, *37*, 1700153.
[12] E. De Clercq, *Biochem. Pharmacol.* **2009**, *77*, 1655–1664.
[13] J. B. Rubin, *Semin. Cancer Biol.* **2009**, *19*, 116–122.
[14] A. Sachpatzidis, B. K. Benton, J. P. Manfredi, H. Wang, A. Hamilton, H. G. Dohlman, E. Lolis, *J. Biol. Chem.* **2003**, *278*, 896–907.
[15] A. Levoye, K. Balabanian, F. Baleux, F. Bachelerie, B. Lagane, *Blood* **2009**, *113*, 6085–6093.
[16] P. M. Murphy, *Annu. Rev. Immunol.* **1994**, *12*, 593–633.
[17] Q. Ma, D. Jones, P. R. Borghesani, R. A. Segal, T. Nagasawa, T. Kishimoto, R. T. Bronson, T. A. Springer, *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 9448–9453.
[18] A. Mantovani, *Immunol. Today* **1999**, *20*, 254–257.
[19] J. A. Burger, T. J. Kipps, *Blood* **2006**, *107*, 1761–1767.
[20] K. S. Kumar, M. Pillong, J. Kunze, I. Burghardt, M. Weller, M. A. Grotzer, G. Schneider, M. Baumgartner, *Sci. Rep.* **2015**, *5*, 15338.
[21] G. W. Bemis, M. A. Murcko, *J. Med. Chem.* **1996**, *39*, 2887–2893.
[22] G. Thoma, C. Beerli, M. Bigaud, C. Bruns, N. G. Cooke, M. B. Streiff, H.-G. Zerwes, *J. Med. Chem.* **2008**, *51*, 7915–7920.
[23] R. T. Skerlj, G. J. Bridger, A. I. Kaller, E. J. McEachern, J. B. Crawford, Y. Zhou, B. Atsma, J. Langille, S. Nan, D. Veale, T. Wilson, C. Harwig, S. Hatse, K. Princen, E. De Clercq, D. Schols, *J. Med. Chem.* **2010**, *53*, 3376–3388.

[24] J. F. Miller, K. S. Gudmundsson, L. D'Aurora Richardson, S. Jenkinson, A. Spaltenstein, M. Thomson, P. Wheelan, *Bioorg. Med. Chem. Lett.* **2010**, *20*, 3026–3030.
[25] R. Skerlj, G. Bridger, E. McEachern, C. Harwig, C. Smith, T. Wilson, D. Veale, H. Yee, J. Crawford, K. Skupinska, R. Wauthy, W. Yang, Y. Zhu, D. Bogucki, M. Di Fluri, J. Langille, D. Huskens, E. De Clercq, D. Schols, *Bioorg. Med. Chem. Lett.* **2011**, *21*, 262–266.
[26] R. Skerlj, G. Bridger, E. McEachern, C. Harwig, C. Smith, A. Kaller, D. Veale, H. Yee, K. Skupinska, R. Wauthy, L. Wang, I. Baird, Y. Zhu, K. Burrage, W. Yang, M. Sartori, D. Huskens, E. De Clercq, D. Schols, *Bioorg. Med. Chem. Lett.* **2011**, *21*, 1414–1418.
[27] V. M. Truax, H. Zhao, B. M. Katzman, A. R. Prosser, A. A. Alcaraz, M. T. Saindane, R. B. Howard, D. Culver, R. F. Arrendale, P. R. Gruddanti, T. J. Evers, M. G. Natchus, J. P. Snyder, D. C. Liotta, L. J. Wilson, *ACS Med. Chem. Lett.* **2013**, *4*, 1025–1030.
[28] J. Zhao, Y. Huang, D. Liu, Y. Chen, *Bioorg. Med. Chem. Lett.* **2015**, *25*, 4950–4955.
[29] H. Zhang, D. Kang, B. Huang, N. Liu, F. Zhao, P. Zhan, X. Liu, *Eur. J. Med. Chem.* **2016**, *114*, 65–78.
[30] D. Das, K. Maeda, Y. Hayashi, N. Gavande, D. V. Desai, S. B. Chang, A. K. Ghosh, H. Mitsuya, *Antimicrob. Agents Chemother.* **2015**, *59*, 1895–1904.
[31] M. Reutlinger, C. P. Koch, D. Reker, N. Todoroff, P. Schneider, T. Rodrigues, G. Schneider, *Mol. Inf.* **2013**, *32*, 133–138.
[32] D. Reker, T. Rodrigues, P. Schneider, G. Schneider, *Proc. Natl. Acad. Sci. USA* **2014**, *111*, 4067–4072.
[33] G. Schneider, P. Schneider, *Expert Opin. Drug Discovery* **2017**, *12*, 271–277.
[34] The Gene Ontology Consortium, *Nucleic Acids Res.* **2017**, *45*, D331–D338.
[35] M. Ashburner, C. A. Ball, J. A. Blake, D. Botstein, H. Butler, J. M. Cherry, A. P. Davis, K. Dolinski, S. S. Dwight, J. T. Eppig, M. A. Harris, D. P. Hill, L. Issel-Tarver, A. Kasarskis, S. Lewis, J. C. Matese, J. E. Richardson, M. Ringwald, G. M. Rubin, G. M. Sherlock, M. Ashburner, *Nat. Genet.* **2000**, *25*, 25–29.
[36] H. Huang, K. Wu, J. Ma, Y. Du, C. Cao, Y. Nie, *Int. Immunopharmacol.* **2016**, *39*, 113–120.
[37] M. M. Mysinger, D. R. Weiss, J. J. Ziarek, S. Gravel, A. K. Doak, J. Karpiak, N. Heveker, B. K. Shoichet, B. F. Volkman, *Proc. Natl. Acad. Sci. USA* **2012**, *109*, 5517–5522.
[38] S. Oishi, T. Kuroyanagi, T. Kubo, N. Montpas, Y. Yoshikawa, R. Misu, Y. Kobayashi, H. Ohno, N. Heveker, T. Furuya, N. Fujii, *J. Med. Chem.* **2015**, *58*, 5218–5225.
[39] Z. G. Zachariassen, S. Thiele, E. A. Berg, P. Rasmussen, T. Fossen, M. M. Rosenkilde, J. Våbenø, B. E. Haug, *Bioorg. Med. Chem.* **2014**, *22*, 4759–4769.
[40] Z. G. Zachariassen, S. Karlshøj, B. E. Haug, M. M. Rosenkilde, J. Våbenø, *J. Med. Chem.* **2015**, *58*, 8141–8153.
[41] H. Ha, B. Debnath, S. Odde, T. Bensman, H. Ho, P. M. Beringer, N. Neamati, *J. Chem. Inf. Model.* **2015**, *55*, 1720–1738.
[42] a) P. Schneider, G. Schneider, *Angew. Chem.* **2017**, *129*, 11678–11682; *Angew. Chem. Int. Ed.* **2017**, *56*, 11520–11524; b) P. Schneider, G. Schneider, *Angew. Chem. Int. Ed.* **2017**, *56*, 11520–11524; *Angew. Chem.* **2017**, *129*, 11678–11682.

Manuscript received: July 2, 2019
Revised manuscript received: September 11, 2019